

WOODHOUSE EXHIBIT 10

EXHIBIT I

Message

From: Chris Marra [REDACTED]@meta.com]
Sent: 5/10/2023 11:57:16 PM
To: Sergey Edunov [REDACTED]@meta.com]; Melanie Kambadur [REDACTED]@meta.com]; Chris Marra [REDACTED]@meta.com]; Logan Kerr [REDACTED]@meta.com]
Subject: Message summary [{"otherUserFbId":null,"threadFbId":6080302042079522}]
Attachments: _A_C_Priv_LLaMa__A_New_Industry_Platform.pdf

Chris Marra (5/10/2023 16:34:47 PDT):

>Hello all! I think everyone has the context here, since Melanie was on the OG LLaMa OSS thread, Sergey was in my meeting today, and I just chatted with Logan, but tl;dr is we're trying to lay out the risk/effort spectrum for mitigations to the OG LLaMa model we /could/ apply before fully releasing it to the community, so we can make a decision on the timing for those. I figured a thread to enumerate the options without that broad XFN might be a little easier.

>

>Logan helped me through the spectrum of:

>1. OG model, no tuning, available today.

>2. OG model, some safety measurement + some tuning, ??? on timing and said to ping Melanie for that and what safety measurements we would do / what datasets would go into the tuning. There's also an in between after this with more mitigations.

>3. v2 model, with new IP/PII mitigations, some time in July - clear wait, but lots of benefit.

>

>Melanie/Sergey, can you help elaborate on #2? I think that's the trade-off we probably have to put in front of Zuck and Ahmad, and understanding how much safety measurement / RHLF we already have planned on the OG model (or is this duplicative work?) on what timeframe can help tee up what we're supposed to get back with by the end of the week.

Melanie Kambadur (5/10/2023 16:37:19 PDT):

>I can help with this in the next day or two, but what timeline seems generally acceptable?

Melanie Kambadur (5/10/2023 16:37:35 PDT):

>and are we planning to do a tuned-model release or only the base pretrained model?

Chris Marra (5/10/2023 16:38:54 PDT):

>that's what we're trying to decide, essentially is the effort to do/wait for a tuned-model enough of an improvement to justify waiting vs. just the base pretrained model

Chris Marra (5/10/2023 16:39:21 PDT):

>The guidance is to not slow down actual productionization work for this, so try to find the optimal path through

Melanie Kambadur (5/10/2023 16:48:33 PDT):

>Ok I see a lot of risks with doing an untuned, relatively unmeasured commercial release from a legal perspective, policy perspective, and researcher/competitor backlash perspective. Not sure to what degree these have already been surfaced by various other people in this effort.

Chris Marra (5/10/2023 16:50:54 PDT):

>Totally, what we're trying to get to Mark and Ahmad by the end of the week is a view on what the lowest hanging fruit improvement on that is - e.g. if we were willing to wait 1 month, where would we be on the journey towards v2 with IP + PII mitigations, and how much better would that be, and if we were willing to wait 2 months, how much better would that be?

Melanie Kambadur (5/10/2023 16:52:47 PDT):

>But has someone put together a list of the risks of doing the low hanging fruit option yet or is that part of what we need to convey here

Chris Marra (5/10/2023 16:56:40 PDT):

>the outcome from Monday reviewing this doc was "Work w/ legal to

Redacted

Redacted

Chris Marra (5/10/2023 16:56:42 PDT):

shared: _A_C_Priv_LLaMa__A_New_Industry_Platform.pdf

Chris Marra (5/10/2023 16:57:16 PDT):

>(I know you were on it at some point before they did the access scrub, here's the original link:
[https://docs.google.com/document/\[REDACTED\]](https://docs.google.com/document/[REDACTED])

